ISSN 2347-1921

ORACLE INEQUALITIES FOR LASSO AND DANTZIG SELECTOR IN HIGH-DIMENSIONAL LINEAR REGRESSION

Shiqing Wang1, Wenfang Shi2

College of Mathematics and Information Sciences, North China University of Water Resources and Electric Power, Zhengzhou 450011, China

Email: wangshiqing@ncwu.edu.cn

College of Mathematics and Information Sciences, North China University of Water Resources and Electric Power, Zhengzhou 450011, China

Email: hnxzswf@163.com

ABSTRACT

During the last few years, a great deal attention has been focused on lasso and Dantzig selector in high-dimensional linear regression under a sparsity scenario, that is, when the number of variables can be much larger than the sample size. The authors [4][11][12] derived sparsity oracle inequalities of lasso and Dantzig selector for the prediction risk and bounds on the $\ell_p (1 \le p \le 2)$ estimation loss under a variety of assumptions. In this paper, we take the restricted eigenvalue conditions, compatibility condition and UDP condition for examples to show oracle inequalities about lasso and Dantzig selector for high-dimensional linear regression.

KEYWORDS

lasso; Dantzig selector; oracle inequalities; restricted eigenvalue conditions; compatibility condition; UDP condition

Council for Innovative Research

Peer Review Research Publishing System

Journal: JOURNAL OF ADVANCES IN MATHEMATICS

Vol .9, No.7

www.cirjam.com , editorjam@gmail.com

1. INTRODUCTION

In many modern applications,one has to deal with very large datasets. Regression problems may involve a large number of covariates, possibly larger than the sample size. In this situation, a major issue lies in dimension reduction, which can be performed through the selection of a Small amount of relevant covariates. For this purpose, numerous regression methods has been proposed in the literature, ranging from the classical information criteria such as C_P ,AIC, and BIC to the more recent regularization-based techniques such as the l_1 penalized least square estimator, known as the lasso and Dantzig selector**[12]**. Their popularity might be due to the fact that these techniques are computationally tractable, even for high-dimensional data when the number of covatiates p is large. Besides, there are much regularization schemes for high-dimensional regression, for example NNLS is a very simple and effective regularization technique for a certain class of high-dimensional regression problems**[6].** Consider the high-dimensional linear model where one observes a vector $y \in R^n$ such that

$$y = X\beta + \varepsilon \quad (1)$$

Where $\beta \in R^p$ is an unknown target vector one would like to recover, $X \in R^{n\times p}$ is called design matrix with possibly much fewer rows than columns, $n << p$, and $\varepsilon \in R^n$ is a random error term that contains all perturbations of the experiment.

A standard hypothesis in high-dimensional regression **[14]** requires that one can provide a constant $\lambda_n^0 \in R$, as small as possible, such that

$$\left\|X^T\varepsilon\right\|_\infty \le \lambda_n^0 \quad (2)$$

With an overwhelming probability. In the case of n-multivariate Gaussian distribution, it is known that $\lambda_n^0 = O\left(\sigma_n\sqrt{\log p}\right)$, where $\sigma_n > 0$ denotes the standard deviation of the noise; (see Lemma A.1.[1])

In general terms, we are interested in accurately estimating the target vector β and $X\beta$ from few and corrupted observations. During the past decade, this challenging issue has attracted a lot of attention among the statistical society. In 1996, Tibshirani introduced the lasso(see[13]):

$$\hat{\beta}_L = \arg\min_{\beta\in R^p}\left\{\frac{1}{2}\|y - X\beta\|_2^2 + \lambda_\ell\|\beta\|_1\right\} \quad (3)$$

Where $\lambda_\ell > 0$ is some tuning constant. Two decades later, this estimator continues to play a key role in our understanding of high-dimensional inverse problems. Its popularity may be due to the fact that this estimator is computational feasibility. Recently, Candes and Tao **[12]**have introduced the the Dantzig selector as

$$\hat{\beta}_D = \arg\min_{\beta\in R^p}\|\beta\|_1 \text{ s.t. } \left\|X^T(y - X\beta)\right\|_\infty \le \lambda_d \quad (4)$$

Where $\lambda_d > 0$ is a tuning parameter. It is known that it can be recasted as a linear program. Hence , it is also computationally tractable. Interestingly, both the lasso and the Dantzig selector can be seen as orthogonal projections of 0 into $DC(s) = \left\{\beta \in R^p, \left\|X^T(y - X\beta)\right\|_\infty \le s\right\}$,using an ℓ_1 distance for the Dantzig selector and ℓ_2 distance for the lasso.

So we also can investigate the properties of estimators defined as projections on $DC(s)$ using general distances**[5].** Based on the linear model, we know that these estimators not only rely on the sparsity of unknown vector and the tuning parameter, but also rely on the distribution assumptions of random errors. In recent years, authors has out forward new methods to improve above problems. Such as the method in [2] allows for very weak distribution assumptions and does not require the knowledge of the variance of random errors. The result in **[3]** tells us the lasso prediction works well for any degree of correlations if suitable tuning parameters are chosen.

There are several objectives may be considered by the statistician when we deal with the model given by Equation(1). Usually, we consider two specific objectives: prediction and estimation in the high-dimensional setting.

ISSN 2347-1921

The reconstruction of the signal $X\beta$ is first considered. The quality of the reconstruction with an estimator $\hat{\beta}$ is often measured with the squared error $\left\|X\hat{\beta}-X\beta\right\|_2^2$, namely prediction risk. Another thinking is that the estimator $\hat{\beta}$ is close to β in terms of the ℓ_q distance for $q\geq 1$,namely bounds on the $\ell_p(1\leq p\leq 2)$ estimation loss.

In this paper, under a sparsity case, we introduce some results about lasso and Dantzig selector about prediction risk and estimation loss for high-dimensional linear regression with noiseless observations.

2.DEFINITIONS AND NOTATIONS

Unless stated otherwise, all through this paper we will assume that $\varepsilon\sim N(0,\sigma^2 I_n)$ $\sigma^2>0$. The analysis of regularized regression methods for high-dimensional data usually involves a sparsity assumption on β .Let

$M(\beta)=\sum_{j=1}^{p} I_{(\beta_j\neq 0)}=|J(\beta)|$ denote the number of nonzero coordinates of β, where $I_{\{.\}}$ denotes the indicator function $J(\beta)=\{j\in\{1,\ldots,P\}:\beta_j\neq 0\}$ and $|J|$ denotes the cardinality of J. The value $M(\beta)$ characterizes the sparsity of the vector β. The smaller $M(\beta)$, the sparser β. For a vector $\delta\in R^p$ and a subset $J\subset\{1,\ldots,P\}$, we denote by δ_J the vector in R^p that has the same coordinates as δ on J and zero coordinates on the complement J^c of J For any $q\geq 1$, $a\in R^p$,denote $\|a\|_q^q=\sum_{i=1}^{p}|a_i|^q$; $\|a\|_\infty=\max_{1\leq i\leq p}|a_i|$, the ℓ_q and ℓ_∞ norms, respectively.

3.RESULTS ABOUT LASSO AND DANTZIG SELECTOR UNDER DIFFERENT ASSUMPTIONS

Now we will show some results about lasso and Dantzig selector in high-dimensional linear regression under different assumptions on X .

3.1 X satisfies $RE(s,c_0)$ [4]

Assumption $RE(s,c_0)$ for some integer $1\leq s\leq p$, $c_0>0$, the following condition holds:

$$\kappa(s,c_0)=\min_{\substack{J\subset\{1,\ldots,P\}\\ |J|\leq s}}\ \min_{\substack{\delta\neq 0\\ \|\delta_{J^c}\|_1\leq c_0\|\delta_J\|_1}}\frac{\|X\delta\|_2}{\|\delta_J\|_2}>0 \quad (5)$$

There is an improvement of the restricted eigenvalue condition, it replace the $\|\delta_J\|_2$ with $\|\delta_J\|_1$ ([10])

Result 1 assume X satisfies $RE(s,c_0)$ and $|J(\beta)|\leq s$ consider the lasso defined by (3), we have

$$\left\|\hat{\beta}_L-\beta\right\|_1\leq\frac{2(\lambda_n^0+\lambda_\ell)(1+c_0)^2 s}{\kappa^2(s,c_0)}$$

$$\left\|X\hat{\beta}_L-X\beta\right\|_2\leq\frac{2(\lambda_\ell+\lambda_n^0)(1+c_0)\sqrt{s}}{\kappa(s,c_0)}$$

Proof let $h=\hat{\beta}_L-\beta\in R^p$

By optimality, we have

ISSN 2347-1921

$$\frac{1}{2}\left\|y - X\hat{\beta}_L\right\|_2^2 + \lambda_\ell\left\|\hat{\beta}_L\right\|_1 \le \frac{1}{2}\|y - X\beta\|_2^2 + \lambda_\ell\|\beta\|_1$$

It yield

$$\frac{1}{2}\|Xh\|_2^2 - \left\langle X^T\varepsilon, h\right\rangle + \lambda_\ell\left\|\hat{\beta}_L\right\|_1 \le \lambda_\ell\|\beta\|_1$$

Using (2) we can get $\left\langle X^T\varepsilon, h\right\rangle \le \left\|X^T\varepsilon\right\|_\infty \|h\|_1 \le \lambda_n^0\|h\|_1$

using the triangle inequality $\|\beta\|_1 - \left\|\hat{\beta}_L\right\|_1 \le \left\|\beta - \hat{\beta}_L\right\|_1 = \|h\|_1$

Then $\frac{1}{2}\|Xh\|_2^2 \le \lambda_\ell\|\beta\|_1 - \lambda_\ell\left\|\hat{\beta}_L\right\|_1 + \lambda_n^0\|h\|_1 \le \left(\lambda_n^0 + \lambda_\ell\right)\|h\|_1$

Based on the inequality $\left\|h_{J^C}\right\|_1 \le c_0\|h_J\|_1$

We get $\|h\|_1 = \|h_J\|_1 + \left\|h_{J^C}\right\|_1 \le (1+c_0)\|h_J\|_1 \le (1+c_0)\sqrt{s}\|h_J\|_2$

Here X satisfies $RE(s, c_0)$ so we can get

$$\frac{\|Xh\|_2}{\|h_J\|_2} \ge \kappa(s, c_0) > 0$$

It equals to $\|h_J\|_2 \le \frac{\|Xh\|_2}{\kappa(s, c_0)}$ or $\|Xh\|_2 \ge \kappa(s, c_0)\|h_J\|_2$

So $\frac{1}{2}\|Xh\|_2^2 \le \left(\lambda_n^0 + \lambda_\ell\right)(1+c_0)\sqrt{s}\|h_J\|_2 \le \left(\lambda_n^0 + \lambda_\ell\right)(1+c_0)\sqrt{s}\,\frac{\|Xh\|_2}{\kappa(s, c_0)}$

Or $\frac{1}{2}\kappa^2(s, c_0)\|h_J\|_2^2 \le \frac{1}{2}\|Xh\|_2^2 \le \left(\lambda_n^0 + \lambda_\ell\right)(1+c_0)\sqrt{s}\|h_J\|_2$

It equals that $\|Xh\|_2 \le \frac{2\left(\lambda_n^0 + \lambda_\ell\right)(1+c_0)\sqrt{s}}{\kappa(s, c_0)}$ or $\|h_J\|_2 \le \frac{2\left(\lambda_n^0 + \lambda_\ell\right)(1+c_0)\sqrt{s}}{\kappa^2(s, c_0)}$

So we can get $\|h\|_1 \le (1+c_0)\sqrt{s}\|h_J\|_2 \le \frac{2\left(\lambda_n^0 + \lambda_d\right)(1+c_0)^2 s}{\kappa^2(s, c_0)}$

Result 2 assume X satisfies $RE(s, c_0)$ and $|J(\beta)| \le s$ consider the Dantzig selector defined by (4), we have

$$\left\|\hat{\beta}_D - \beta\right\|_1 \le \frac{\left(\lambda_n^0 + \lambda_d\right)(1+c_0)^2 s}{\kappa^2(s, c_0)}$$

$$\left\|X\hat{\beta}_D - X\beta\right\|_2 \le \frac{\left(\lambda_d + \lambda_n^0\right)(1+c_0)\sqrt{s}}{\kappa(s, c_0)}$$

Proof Let $h = \hat{\beta}_D - \beta \in R^p$

ISSN 2347-1921

Recall that $\left\|X^T\varepsilon\right\|_\infty \le \lambda_n^0$, it yields

$$\|Xh\|_2^2 \le \left\|X^TXh\right\|_\infty \|h\|_1 = \left\|X^T\left(y - X\hat{\beta}_D\right) + X^T(X\beta - y)\right\|_\infty \|h\|_1 \le \left(\lambda_d + \lambda_n^0\right)\|h\|_1$$

Based on the inequality $\left\|h_{J^C}\right\|_1 \le c_0\|h_J\|_1$

We get $\|h\|_1 = \|h_J\|_1 + \left\|h_{J^C}\right\|_1 \le (1+c_0)\|h_J\|_1 \le (1+c_0)\sqrt{s}\|h_J\|_2$

Here X satisfies $RE(s,c_0)$ so we can get

$$\frac{\|Xh\|_2}{\|h_J\|_2} \ge \kappa(s,c_0) > 0$$

It equals to $\|h_J\|_2 \le \dfrac{\|Xh\|_2}{\kappa(s,c_0)}$ or $\|Xh\|_2 \ge \kappa(s,c_0)\|h_J\|_2$

So $\|Xh\|_2^2 \le \left(\lambda_n^0 + \lambda_d\right)(1+c_0)\sqrt{s}\|h_J\|_2 \le \left(\lambda_n^0 + \lambda_d\right)(1+c_0)\sqrt{s}\,\dfrac{\|Xh\|_2}{\kappa(s,c_0)}$

Or $\kappa^2(s,c_0)\|h_J\|_2^2 \le \|Xh\|_2^2 \le \left(\lambda_d^0 + \lambda_\ell\right)(1+c_0)\sqrt{s}\|h_J\|_2$

It equals that $\|Xh\|_2 \le \dfrac{\left(\lambda_d^0 + \lambda_\ell\right)(1+c_0)\sqrt{s}}{\kappa(s,c_0)}$ or $\|h_J\|_2 \le \dfrac{\left(\lambda_n^0 + \lambda_d\right)(1+c_0)\sqrt{s}}{\kappa^2(s,c_0)}$

So we can get $\|h\|_1 \le (1+c_0)\sqrt{s}\|h_J\|_2 \le \dfrac{\left(\lambda_d^0 + \lambda_\ell\right)(1+c_0)^2 s}{\kappa^2(s,c_0)}$

3.2 X satisfies compatibility condition [8]

A matrix $X \in R^{n\times p}$ satisfies the Compatibility condition (S,c_0) if and only if

$$\phi(S,c_0) = \min_{\substack{J\subset\{1,\dots,p\}\\ |J|\le s}} \min_{\substack{\delta\ne 0\\ \left\|h_{J^c}\right\|_1 \le c_0\|h_J\|_1}} \frac{\sqrt{|J|}\|Xh\|_2}{\|h_J\|_1} > 0 \quad (6)$$

Result 1 assume that X satisfies the Compatibility condition (S,c_0) and $|J(\beta)| \le s$ consider the lasso defined by (3), we have

$$\left\|\hat{\beta}_L - \beta\right\|_1 \le \frac{2\left(\lambda_n^0 + \lambda_\ell\right)(1+c_0)^2 s}{\phi^2(s,c_0)}$$

$$\left\|X\hat{\beta}_L - X\beta\right\|_2 \le \frac{2\left(\lambda_\ell + \lambda_n^0\right)(1+c_0)\sqrt{s}}{\phi(s,c_0)}$$

Proof let $h = \hat{\beta}_L - \beta \in R^p$

By optimality, we have

$$\frac{1}{2}\left\|y-X\hat{\beta}_L\right\|_2^2+\lambda_\ell\left\|\hat{\beta}_L\right\|_1 \le \frac{1}{2}\|y-X\beta\|_2^2+\lambda_\ell\|\beta\|_1$$

It yield

$$\frac{1}{2}\|Xh\|_2^2-\left\langle X^T\varepsilon,h\right\rangle+\lambda_\ell\left\|\hat{\beta}_L\right\|_1 \le \lambda_\ell\|\beta\|_1$$

Using (2) we can get $\left\langle X^T\varepsilon,h\right\rangle \le \left\|X^T\varepsilon\right\|_\infty\|h\|_1 \le \lambda_n^0\|h\|_1$

using the triangle inequality $\|\beta\|_1-\left\|\hat{\beta}_L\right\|_1 \le \left\|\beta-\hat{\beta}_L\right\|_1=\|h\|_1$

Then $\frac{1}{2}\|Xh\|_2^2 \le \lambda_\ell\|\beta\|_1-\lambda_\ell\left\|\hat{\beta}_L\right\|_1+\lambda_n^0\|h\|_1 \le \left(\lambda_n^0+\lambda_\ell\right)\|h\|_1$

Based on the inequality $\left\|h_{J^C}\right\|_1 \le c_0\|h_J\|_1$

We get $\|h\|_1=\|h_J\|_1+\left\|h_{J^C}\right\|_1 \le (1+c_0)\|h_J\|_1$

Here X satisfies Compatibility condition (S,c_0) so we can get

$$\frac{\sqrt{|J|}\|Xh\|_2}{\|h_J\|_1} \ge \phi(S,c_0)>0$$

It equals to $\|h_J\|_1 \le \frac{\sqrt{|J|}\|Xh\|_2}{\phi(S,c_0)} \le \frac{\sqrt{s}\|Xh\|_2}{\phi(S,c_0)}$ or

$$\|Xh\|_2 \ge \frac{\phi(s,c_0)\|h_J\|_1}{\sqrt{|J|}} \ge \frac{\phi(s,c_0)\|h_J\|_1}{\sqrt{s}}$$

So $\frac{1}{2}\|Xh\|_2^2 \le \left(\lambda_n^0+\lambda_\ell\right)(1+c_0)\|h_J\|_1 \le \left(\lambda_n^0+\lambda_\ell\right)(1+c_0)\frac{\sqrt{s}\|Xh\|_2}{\phi(S,c_0)}$

Or $\frac{1}{2}\frac{\phi^2(s,c_0)}{s}\|h_J\|_1^2 \le \frac{1}{2}\|Xh\|_2^2 \le \left(\lambda_n^0+\lambda_\ell\right)(1+c_0)\|h_J\|_1$

It equals that $\|Xh\|_2 \le \frac{2\left(\lambda_n^0+\lambda_\ell\right)(1+c_0)\sqrt{s}}{\kappa(s,c_0)}$ or $\|h_J\|_1 \le \frac{2\left(\lambda_n^0+\lambda_\ell\right)(1+c_0)^2 s}{\phi^2(s,c_0)}$

So we can get $\|h\|_1 \le (1+c_0)\|h_J\|_1 \le \frac{2\left(\lambda_n^0+\lambda_\ell\right)(1+c_0)^2 s}{\phi^2(s,c_0)}$

Result 2 assume X satisfies Compatibility condition $(S,c_0)$and $|J(\beta)| \le s$ consider the Dantzigselector defined by (4), we have

$$\left\|X\hat{\beta}_D-X\beta\right\|_2 \le \frac{\left(\lambda_d+\lambda_n^0\right)(1+c_0)\sqrt{s}}{\phi(s,c_0)}$$

$$\left\|\hat{\beta}_D-\beta\right\|_1 \le \frac{(\lambda_n^0+\lambda_d)(1+c_0)^2 s}{\phi^2(s,c_0)}$$

Proof Let $h=\hat{\beta}_D-\beta\in R^p$

Recall that $\left\|X^T\varepsilon\right\|_\infty \le \lambda_n^0$, it yields

$$\|Xh\|_2^2 \le \left\|X^TXh\right\|_\infty \|h\|_1 = \left\|X^T\left(y-X\hat{\beta}_D\right)+X^T(X\beta-y)\right\|_\infty \|h\|_1 \le \left(\lambda_d+\lambda_n^0\right)\|h\|_1$$

Based on the inequality $\left\|h_{J^C}\right\|_1 \le c_0\|h_J\|_1$

We get $\|h\|_1=\|h_J\|_1+\left\|h_{J^C}\right\|_1 \le (1+c_0)\|h_J\|_1$

Here X satisfies Compatibility condition (S,c_0) so we can get

$$\frac{\sqrt{|J|}\|Xh\|_2}{\|h_J\|_1} \ge \phi(S,c_0) > 0$$

It equals to $\|h_J\|_1 \le \dfrac{\sqrt{|J|}\|Xh\|_2}{\phi(S,c_0)} \le \dfrac{\sqrt{s}\|Xh\|_2}{\phi(S,c_0)}$

or $\|Xh\|_2 \ge \dfrac{\phi(s,c_0)\|h_J\|_1}{\sqrt{|J|}} \ge \dfrac{\phi(s,c_0)\|h_J\|_1}{\sqrt{s}}$

So $\|Xh\|_2^2 \le (\lambda_n^0+\lambda_d)(1+c_0)\|h_J\|_1 \le (\lambda_n^0+\lambda_d)(1+c_0)\sqrt{s}\dfrac{\|Xh\|_2}{\phi(s,c_0)}$

Or $\dfrac{\phi^2(s,c_0)\|h_J\|_1^2}{s} \le \|Xh\|_2^2 \le (\lambda_d^0+\lambda_\ell)(1+c_0)\|h_J\|_1$

It equals that $\|Xh\|_2 \le \dfrac{(\lambda_n^0+\lambda_d)(1+c_0)\sqrt{s}}{\phi(s,c_0)}$ or $\|h_J\|_1 \le \dfrac{(\lambda_n^0+\lambda_d)(1+c_0)s}{\phi^2(s,c_0)}$

So we can get $\|h\|_1 \le (1+c_0)\|h_J\|_1 \le \dfrac{(\lambda_n^0+\lambda_d)(1+c_0)^2 s}{\phi^2(s,c_0)}$

based on the definition of Compatibility condition (S,c_0) and the restricted eigenvalue condition, we can deduce that $\phi(s,c_0)\ge RE(s,c_0)$, so

$$\|Xh\|_2 \le \frac{(\lambda_n^0+\lambda_d)(1+c_0)\sqrt{s}}{\phi(s,c_0)} \le \frac{(\lambda_n^0+\lambda_d)(1+c_0)\sqrt{s}}{\kappa(s,c_0)}$$

$$\|h_J\|_1 \le \frac{(\lambda_n^0+\lambda_d)(1+c_0)s}{\phi^2(s,c_0)} \le \frac{(\lambda_n^0+\lambda_d)(1+c_0)s}{\kappa^2(s,c_0)}$$

we can proof that the compatibility condition is the weaker than the restricted eigenvalue condition.(see[9])

3.3 X satisfies $UDP(S_0,\kappa_0,\Delta)$ [1]

$UDP(S_0,\kappa_0,\Delta)$ Given $1 \le S_0 \le P$ and $0 < \kappa_0 < \frac{1}{2}$, we say that a matrix $X \in R^{n\times p}$ satisfies the universal distortion condition of order S_0, magnitude κ_0 and parameter Δ if and only if for all $\delta \in R^p$, for all intergers $s \in \{1,\ldots,S_0\}$, for all subsets $J \subseteq \{1,\ldots,p\}$ such that $|J| = s$, it holds

$$\|\delta_J\|_1 \le \Delta\sqrt{s}\|X\delta\|_2 + \kappa_0\|\delta\|_1 \quad (7)$$

Result1: assume that X satisfies $UDP(S_0,\kappa_0,\Delta)$ with $\kappa_0 < \frac{1}{2}$ and that (2) holds. Then for any $\lambda_\ell > \lambda_n^0/1-2\kappa_0$, it holds

$$\left\|\hat{\beta}_L - \beta\right\|_1 \le \frac{2}{\left(1-\frac{\lambda_n^0}{\lambda_\ell}\right)-2\kappa_0} \min_{\substack{J\subseteq\{1,\ldots,p\}\\ |J|=s,s\le S_0}} \left(\lambda_\ell\Delta^2 s + \left\|\beta_{J^c}\right\|_1\right)$$

$$\left\|X\hat{\beta}_L - X\beta\right\|_2 \le \min_{\substack{J\subseteq\{1,\ldots,p\}\\ |J|=s,s\le S_0}} \left(4\lambda_\ell\Delta\sqrt{s} + \frac{\left\|\beta_{J^c}\right\|_1}{\Delta\sqrt{s}}\right)$$

See theorem 2.1, 2.2.[1]

Proof. Let $h = \hat{\beta}_L - \beta \in R^p$ and $\lambda_\ell \ge \lambda_n^0$

by optimality, we have

$$\frac{1}{2}\left\|X\hat{\beta}_L - y\right\|_2^2 + \lambda_\ell\left\|\hat{\beta}_L\right\|_1 \le \frac{1}{2}\|X\beta - y\|_2^2 + \lambda_\ell\|\beta\|_1$$

It yields

$$\frac{1}{2}\|Xh\|_2^2 - \left\langle X^T\varepsilon, h\right\rangle + \lambda_\ell\left\|\hat{\beta}_L\right\|_1 \le \lambda_\ell\|\beta\|_1$$

Let $J \subseteq \{1,\ldots,p\}$; we have

$$\begin{aligned}\frac{1}{2}\|Xh\|_2^2 + \lambda_\ell\left\|\hat{\beta}_{L_{J^c}}\right\|_1 &\le \lambda_\ell\left(\|\beta_J\|_1 - \left\|\hat{\beta}_{L_J}\right\|_1\right) + \lambda_\ell\left\|\beta_{J^c}\right\|_1 + \left\langle X^T\varepsilon, h\right\rangle \\ &\le \lambda_\ell\|h_J\|_1 + \lambda_\ell\left\|\beta_{J^c}\right\|_1 + \lambda_n^0\|h\|_1\end{aligned}$$

Using(2). Adding $\lambda_\ell\left\|\beta_{J^c}\right\|_1$ on both sides, it holds

$$\frac{1}{2}\|Xh\|_2^2 + \left(\lambda_\ell - \lambda_n^0\right)\left\|h_{J^c}\right\|_1 \le \left(\lambda_\ell + \lambda_n^0\right)\left\|h_J\right\|_1 + 2\lambda_\ell\left\|\beta_{J^c}\right\|_1$$

Adding $\left(\lambda_\ell - \lambda_n^0\right)\left\|h_J\right\|_1$ on both sides, we can get

$$\frac{1}{2}\|Xh\|_2^2 + \left(\lambda_\ell - \lambda_n^0\right)\|h\|_1 \le 2\lambda_\ell\left\|h_J\right\|_1 + 2\lambda_\ell\left\|\beta_{J^c}\right\|_1$$

Using（7）, it follows that

$$\frac{1}{2}\|Xh\|_2^2+\left(\lambda_\ell-\lambda_n^0\right)\|h\|_1\le 2\lambda_\ell\left(\Delta\sqrt{s}\|Xh\|_2+\kappa_0\|h\|_1\right)+2\lambda_\ell\left\|\beta_{J^c}\right\|_1$$

$$\frac{1}{2\lambda_\ell}\left[\frac{1}{2}\|Xh\|_2^2+\left(\lambda_\ell-\lambda_n^0\right)\|h\|_1\right]\le\Delta\sqrt{s}\|Xh\|_2+\kappa_0\|h\|_1+\left\|\beta_{J^c}\right\|_1$$

It yields that

$$\left[\frac{1}{2}\left(1-\frac{\lambda_\ell}{\lambda_n^0}\right)-\kappa_0\right]\|h\|_1\le\left(-\frac{1}{4\lambda_\ell}\|Xh\|_2^2+\Delta\sqrt{s}\|Xh\|_2\right)+\left\|\beta_{J^c}\right\|_1$$
$$\le\lambda_\ell\Delta^2 s+\left\|\beta_{J^c}\right\|_1$$

Using the fact that $2ab-b^2\le a^2$, this concludes the proof.

Using this equality $\frac{1}{2\lambda_\ell}\left[\frac{1}{2}\|Xh\|_2^2+\left(\lambda_\ell-\lambda_n^0\right)\|h\|_1\right]\le\Delta\sqrt{s}\|Xh\|_2+\kappa_0\|h\|_1+\left\|\beta_{J^c}\right\|_1$

And $\lambda_\ell>\lambda_n^0/1-2\kappa_0$ we can get

$$\|Xh\|_2^2-4\lambda_\ell\Delta\sqrt{s}\|Xh\|_2\le 4\lambda_\ell\left\|\beta_{J^c}\right\|_1$$

This latter is of the form $x^2-bx\le c$ which implies that $x\le b+c/b$. Hence,

$$\|Xh\|_2\le 4\lambda_\ell\Delta\sqrt{s}+\frac{\left\|\beta_{J^c}\right\|_1}{\Delta\sqrt{s}}$$

Result 2: assume that X satisfies $UDP(S_0,\kappa_0,\Delta)$ with $\kappa_0<\frac{1}{4}$ and that (2) holds. Then for any $\lambda_d>\lambda_n^0/1-4\kappa_0$, it holds

$$\left\|\hat{\beta}_D-\beta\right\|_1\le\frac{4}{\left(1-\frac{\lambda_n^0}{\lambda_d}\right)-4\kappa_0}\min_{\substack{J\subseteq\{1,\ldots,p\}\\|J|=s,s\le S_0}}\left(\lambda_d\Delta^2 s+\left\|\beta_{J^c}\right\|_1\right)$$

$$\left\|X\hat{\beta}_D-X\beta\right\|_2\le\min_{\substack{J\subseteq\{1,\ldots,p\}\\|J|=s,s\le S_0}}\left(4\lambda_d\Delta\sqrt{s}+\frac{\left\|\beta_{J^c}\right\|_1}{\Delta\sqrt{s}}\right)$$

See theorem 2.3, 2.4.[1]

Proof Let $h=\hat{\beta}_D-\beta\in R^p$, $\lambda_\ell\ge\lambda_n^0$ and $J\subseteq\{1,\ldots,p\}$

Recall that $\left\|X^T\varepsilon\right\|_\infty\le\lambda_n^0$, it yields

$$\|Xh\|_2^2\le\left\|X^TXh\right\|_\infty\|h\|_1=\left\|X^T\left(y-X\hat{\beta}_D\right)+X^T(X\beta-y)\right\|_\infty\|h\|_1$$
$$\le\left(\lambda_d+\lambda_n^0\right)\|h\|_1$$

Hence we get

$$\|Xh\|_2^2-(\lambda_d+\lambda_n^0)\|h_{J^c}\|_1\le(\lambda_d+\lambda_n^0)\|h_J\|_1$$

Since $\hat{\beta}_D$ is feasible, it yields $\left\|\hat{\beta}_D\right\|_1\le\|\beta\|_1$. Thus

$$\left\|\hat{\beta}_{D_{J^c}}\right\|_1\le\|\beta_J\|_1-\left\|\hat{\beta}_J\right\|_1+\|\beta_{J^c}\|_1\le\|h_J\|_1+\|\beta_{J^c}\|_1$$

Since $\|h_{J^c}\|_1\le\left\|\hat{\beta}_{D_{J^c}}\right\|_1+\|\beta_{J^c}\|_1$, it yields

$$\|h_{J^c}\|_1\le\|h_J\|_1+2\|\beta_{J^c}\|_1$$

Multiply by $2\lambda_d$ to both sides we can get

$$2\lambda_d\|h_{J^c}\|_1\le2\lambda_d\|h_J\|_1+4\lambda_d\|\beta_{J^c}\|_1$$

Combing this inequality and $\|Xh\|_2^2-(\lambda_d+\lambda_n^0)\|h_{J^c}\|_1\le(\lambda_d+\lambda_n^0)\|h_J\|_1$ we get

$$\|Xh\|_2^2+(\lambda_d-\lambda_n^0)\|h_{J^c}\|_1\le(3\lambda_d+\lambda_n^0)\|h_J\|_1+4\lambda_d\|\beta_{J^c}\|_1$$

Adding $(\lambda_d-\lambda_n^0)\|h_J\|_1$ on both sides, we get

$$\|Xh\|_2^2+(\lambda_d-\lambda_n^0)\|h\|_1\le4\lambda_d\|h_J\|_1+4\lambda_d\|\beta_{J^c}\|_1$$

Here X satisfies $UDP(S_0,\kappa_0,\Delta)$, namely $\|h_J\|_1\le\Delta\sqrt{s}\|Xh\|_2+\kappa_0\|h\|_1$ so

$$\|Xh\|_2^2+(\lambda_d-\lambda_n^0)\|h\|_1\le4\lambda_d\Delta\sqrt{s}\|Xh\|_2+4\lambda_d\kappa_0\|h\|_1+4\lambda_d\|\beta_{J^c}\|_1$$

Thus $\frac{1}{4\lambda_d}\left[\|Xh\|_2^2+(\lambda_d-\lambda_n^0)\|h\|_1\right]\le\Delta\sqrt{s}\|Xh\|_2+\kappa_0\|h\|_1+\|\beta_{J^c}\|_1$

It yields,

$$\begin{aligned}\left[\frac{1}{4}\left(1-\frac{\lambda_n^0}{\lambda_d}\right)-\kappa_0\right]\|h\|_1&\le\left(-\frac{1}{4\lambda_d}\|Xh\|_2^2+\Delta\sqrt{s}\|Xh\|_2\right)+\|\beta_{J^c}\|_1\\&\le\lambda_d\Delta^2s+\|\beta_{J^c}\|_1\end{aligned}$$

Using the fact that $2ab-b^2\le a^2$, this concludes the proof.

$$\|h\|_1\le\frac{\lambda_d\Delta^2s+\|\beta_{J^c}\|_1}{\left[\frac{1}{4}\left(1-\frac{\lambda_n^0}{\lambda_d}\right)-\kappa_0\right]}\le\frac{4}{\left(1-\frac{\lambda_n^0}{\lambda_d}\right)-4\kappa_0}\min_{\substack{J\subseteq\{1,\ldots,p\}\\|J|=s,s\le S_0}}\left(\lambda_d\Delta^2s+\|\beta_{J^c}\|_1\right)$$

Using this equality $\|Xh\|_2^2+(\lambda_d-\lambda_n^0)\|h\|_1\le4\lambda_d\Delta\sqrt{s}\|Xh\|_2+4\lambda_d\kappa_0\|h\|_1+4\lambda_d\|\beta_{J^c}\|_1$ and $\lambda_d>\lambda_n^0/1-4\kappa_0$,we can get

$$\|Xh\|_2^2-4\lambda_\ell\Delta\sqrt{s}\|Xh\|_2\le4\lambda_\ell\|\beta_{J^c}\|_1$$

This latter is of the form $x^2 - bx \le c$ which implies that $x \le b + c/b$. Hence,

$$\|Xh\|_2 \le 4\lambda_d \Delta\sqrt{s} + \frac{\|\beta_{J^c}\|_1}{\Delta\sqrt{s}}$$

4.SUMMARY

Oracle inequalities for lasso and Dantzig selector in linear models have been established under a variety of different assumptions on the design matrix.and how the different conditions and concepts relate to see other **[9].**We know that the restricted eigenvalue conditions or the slightly weaker compatibility condition are sufficient for oracle results. In this paper, we not only show these oracle results under eigenvalue conditions, and compatibility condition but also show results under UDP condition. We also proof that compatibility condition is weaker than the restricted eigenvalue conditions. The UDP condition is similar to the them, see proposition 3.1.**[1]**. As a matter of fact, the UDP condition, the restricted eigenvalue conditions, and compatibility condition are expressions with the same flavor: they aim at controlling the eigenvalues of X on a cone $\left\{\delta \in R^p \middle| \forall s \in \{1,\ldots,p\}, s.t. |\delta| \le s, \|\delta_{J^c}\|_1 \le c\|\delta_J\|_1\right\}$, where $c > 0$ is a tuning parameter.

REFERENCES

[1] Yohann de Castro.2012. A remark on the lasso and Dantzig selector. Statistics and probability letters 83(2013) 304-314.

[2] Eric Gautier., Alexander B. Tsybakov. 2013. Pivotal estimation in high-dimensional regression via linear programming. arXiv:1303.7092v2.

[3] Mohamed Hebiri, Johannes C. Lederer. 2012. How correlations influence lasso prediction. arXiv:1204.1605v2.

[4] Bickel, P., Ritov, Y., Tsybakov A., 2009. Simultaneous analysis of lasso and Dantzig selector. Annals of Statistics37(4), 1705-1732.

[5] Pierre Alquier., Mohamed Hebiri. 2011. Generalization of ℓ_1 constraints for high dimensional regression problems. Statistics and probability letters 81(2011) 1760-1765.

[6] Nicolai Meinshausen. 2012. Sign-constraint least squares estimation for high-dimensional regression. arXiv:1202.0889v1.

[7] Edgar Dobriban., Jianqing Fan.2013. Regularity properties of high-dimensional covariate matrices. arXiv:1305.5198v1.

[8] Van de Geer, S. 2007. The deterministic lasso. In JSM proceedings, (see also http:// stat. Ethz.ch/research/research-reports/2007/140). American Statistical Association.

[9] Van de Geer, S. Bűhlmann, P., 2009. On the conditions used to prove oracle results for the lasso. Electron. J. Stat.3, 1360-1392.

[10] Shiqing Wang., Limin Su. 2013. The oracle inequalities on simulatenous lasso and Dantzig selector in high-dimensional nonparametric regression. Mathematical problems in engieering. Article ID 571361,6pages

[11] Van de Geer, S. (2008). High-dimensional generalized linear models and the lasso. Annals of Statistics 36 614-645.

[12] Candes, E. and Tao, T. 2007. The Dantzig selector: statistical estimation when p is much larger than n. Annals of Statistics 35 2313-2351.

[13] Tibshirani, R., 1996. Regression shrinkage and selection via the lasso.J.R.Stat.Soc.Ser.B 58(1), 267-288

[14]. Hastie, T., Tibshirani, R., Friedman, J., 2009. high-dimensional problems: pn. In: the elements of statistical learing. pp. 1-50

Author's biography

My advisor is Shiqing Wang. He is a Professor of Mathematics at North China University of Water Resources and Electric Power, Zhengzhou, China. His field of interest is Mathematics, Information Sciences and Statistics. He has more than 80 research articles in journal of Mathematics, Information Sciences and Statistics.

I am a Master' of Applied Mathematics, at College of Mathematics and Information Sciences, North China University of Water Resources and Electric Power. My advisor and I study in the same direction. The following is the paper he published in recent two years.

ARTICLE

[1] Shiqing Wang, Limin Su. A mistake on the computation of Jacobians of singular random matrices. International Journal of Applied Mathematics & Statistics. 2013, 40, 10: 1-12. ISSN: 0973-1377. (EI)

[2] Shiqing Wang, Limin Su. Simultaneous Lasso and Dantzig selector in high dimensional nonparametric regression. International Journal of Applied Mathematics & Statistics. 2013, 42, 12:103-118.ISSN: 0973-1377.(EI)

[3] Shiqing Wang, Limin Su. The oracle inequalities on simultaneous Lasso and Dantzig selector in high-dimensional nonparametric regression. Mathematical Problems in Engineering vol.2013, doi:10.1155/2013/571361. ISSN: 1024-123X (Print), ISSN: 1563-5147 (Online). (SCI)

[4] Shiqing Wang, Limin Su. Recovery of high-dimensional sparse signals via L1-minimization. Journal of Applied Mathematics, vol. 2013, doi:10.1155/2013/636094. ISSN: 1110-757X (Print),ISSN: 1687-0042 (Online). (SCI)

[5] Shiqing Wang, Limin Su. New bounds of mutual incoherence property on sparse signals recovery. International Journal of Applied Mathematics & Statistics. 2013, 47, 17: 462-477. ISSN: 0973-1377. (EI)

[6] Shiqing Wang, Yan Shi, and Limin Su. Sparsity oracle inequalities for Lasso and Dantzig selector in high-dimensional nonparametric regression. International Journal of Applied Mathematics & Statistics. 2013, 51, 21: 109-120. ISSN: 0973-1377. (EI)

[7] Shiqing Wang, Yan Shi, and Limin Su. Weaker regularity conditions and sparse recovery in high-dimensional regression . Journal of Applied Mathematics, vol. 2014, doi:10.1155/2014/946241. ISSN: 1110-757X (Print), ISSN: 1687-0042 (Online). (SCI)

[8] Shiqing Wang, Yan Shi, and Limin Su. The hyperbolic region for restricted isometry constants in compressed sensing. International Journal of Circuits, Systems and Signal Processing, 2014, 8: ISSN: 1998-4464. (EI)

[9] Shiqing Wang, Limin Su. Some bounds and the conditional maximum bound for restricted isometry constants. GAZI University Journal of Science, 2014, 27, 4: ISSN: 2147-1762. (EI)

[10] Shiqing Wang, Yan Shi, and Limin Su. Revised bounds on restricted isometry constants in compressed sensing with correction to some reported results. ICIC Express Letters, 2015, 9, 1: ISSN: 1881-803X. (EI)